基于 SSU-SGD 的动态手持物体识别 *

赵文仓，陈聪聪，郑鸿磊

(青岛科技大学 自动化与电子工程学院，山东 青岛 266061)

摘　要：连续目标中包含更加丰富的信息，为了更好的获取动态手持物体中的视觉信息，以不同背景下的动态手持物体为目标，基于步长自学习更新的 SGD 算法（简称 SSU-SGD）提出了适用于动态手持物体识别的三个基准，通过自学习出不同的步长，分别在已知类、未知类和已知对象的基础上进行巩固训练，用于后续的动态手持物体识别中。用 Alexnet 网络和 VGG 网络对三个不同基准下的 naive 策略和累积策略进行了编程实验与仿真，经实验验证，该方法可以有效提高运行速度和训练精确度，并且有效提高了动态手持物体识别过程的实时性，可以进一步的应用于实际。

关键词：连续目标；SSU-SGD 算法；动态手持物体识别的基准

中图分类号：TP391.41　　doi: 10.19734/j.issn.1001-3695.2018.05.0568

Dynamic handheld object recognition based on SSU-SGD

Zhao Wencang, Chen Congcong, Zheng Honglei

(*Institute of Automation & Electronic Engineering*, *Qingdao University of Science & Technology*, *Qingdao Shandong* 266061, *China*)

Abstract: Continuous object contains richer information. In order to obtain more meaningful visual information in dynamic handheld objects, targeting dynamic hand-held objects in different backgrounds. Based on the Step-size Self-learning Update SGD（SSU-SGD）, proposing three benchmarks for dynamic hand-held object recognition . Through self-learning of different step sizes, consolidating the train on the basis of known classes, unknown classes, and known objects for subsequent dynamic hand-held object recognition. Using Alexnet and VGG networks, performing the programming experiments and simulations of naive and cumulative strategies under the three different benchmarks. After experimental verification, this method can effectively improve the running speed and training accuracy and effectively improve the real-time nature of the dynamic hand-held object recognition process can be further applied to the actual.

Key words: continuous object; SSU-SGD; dynamic hand-held object recognition

0　引言

连续目标在日常生活中很常见，如物体的移动、翻转[1]，连续目标往往可以展现更丰富的信息量，通过分析连续多帧图像，可以获得更多有意义的视觉信息[2]。人眼不仅可以看到静止的物体，还可以对运动的目标进行识别、定位和跟踪[3, 4]，从图像识别[5]到视频检测技术[6]，连续目标识别也是人工智能领域的研究热点。

目前，单帧图像识别领域的技术已经比较完善，然而对于高维数据流的识别是一项值得挑战的研究[7]，由于高维数据流的样本库非常庞大，识别方法也就更加复杂[8]。深度学习方法成功的应用于图像识别[9]、语音识别[10]和自然语音处理[11]等领域，同理深度学习方法也可较好的应用于动态手持物体识别。

本文利用深度学习[12, 13]方法对动态的手持物体[14]进行识别，选用 Alexnet 网络和 VGG 网络训练，Alexnet 网络[15]在图像识别领域已取得较为优异的成果，但是在连续目标识别领域的精确度不高；而 VGG 网络[16]的最大特点就是卷积层多，计算量大，但是训练速度较慢。本文提出的 SSU-SGD 算法可以根据特征出现的频率的不同使得相应参数自学出不同的步长，这种方法可以避免 SGD 算法[17, 18]中调整步长的麻烦，提高迭代效率，同时可以避免步长不断减小导致迭代停止，并且在此基础上提出了三个适用于动态手持物体识别的基准，该基准可以有效的提高识别的精确度，快速得到动态手持物体识别问题的最优解，可以更好的把深度学习的方法应用到高维数据流的识别领域。实验表明这种基于 SSU-SGD 算法的动态手持物体识别基准可以有效的提高训练速度和分类精确度。

1　SSU-SGD 算法

1.1　SGD

梯度下降法（SGD）是当下深度学习领域较为流行的优化算法，梯度下降法大致分为三种：需要整个训练集的批量梯度下降、仅用部分样本训练的小批量梯度下降和仅用一个随机样本进行优化的随机梯度下降。

SGD 是一种通过在目标函数梯度 $\nabla_{\lambda}J(\lambda)$ 的反向上更新模型参数，使得模型参数的目标函数 $J(\lambda)$ 最小化的方法。以一个训练样本 X_t 和标签 Y_t 进行一次参数更新， SGD 的迭代公式如下：

$$\lambda=\lambda-\nabla_{\lambda}J(\lambda,x^i,y^i) \tag{1}$$

其中：

$$\nabla_{\lambda}=(y^i-h_{\lambda}(x^i))\,x^i_j \tag{2}$$

收稿日期：2018-05-23；修回日期：2018-07-04　　基金项目：国家自然科学基金资助项目（61171131）；山东省重点研发计划资助项目（2013YD01033）

作者简介：赵文仓（1973-），男，山东青岛人，教授，博士，主要研究方向为智能科学与技术（zhao_coinslab@outlook.com）；陈聪聪（1994-），女，山东潍坊人，硕士，主要研究方向为模式识别与智能系统；郑鸿磊（1995-），男，山东临沂人，硕士，主要研究方向为图像处理与识别.

λ 表示迭代的权重系数，x^i 和 y^i 分别表示第 i 组的输入特征和输出特征，$h_\lambda(x^i)$ 表示拟合的参数，对于一系列步长 η_k，SGD 采用：

$$\lambda_{k+1} = \lambda_k - \eta_k f^{'}(\lambda_k) \tag{3}$$

由于 SGD 通过每次计算只选择一个样本，因此，SGD 速度会更快。

但在梯度下降的同时会使得目标函数值产生巨大的波动，即方差较大[19, 20]。SGD 的另一个关键点是步长的选择问题，步长选择过大或过小，都会导致无法正常收敛[21]。

1.2　SSU-SGD 算法的实现

为了避免 SGD 算法中调整步长的麻烦，提高迭代效率，本文提出的 SSU-SGD 算法，这种自学习出不同步长的 SGD 算法既可以避免步长的不断减小使得参数不再自动调整，终止训练过程；而且可以解决由于步长设置不当而跳过最优解甚至无法得到最优解的问题。

它的主要思想根据特征出现的频率不同使得相应的参数可以自学习出不同的步长，即使得特征出现频率高的参数自学习到较低的步长，特征出现频率低的参数自学习到较低的步长，给出两个步长自学习的法则，其中：

法则 1　在迭代更新频率高时，自学习出较小的步长：

$$\eta \leftarrow R^{(n_i/\omega)} * (\eta_0 / \gamma^{n/\omega_0}) + \varepsilon \tag{4}$$

法则 2　在迭代更新频率低时，自学习出较大的步长：

$$\eta \leftarrow R^{(n_i/\omega)} * \eta_0 + \varepsilon \tag{5}$$

其中：R 表示每次更新的幅度，ω 表示迭代更新的频率，在迭代更新频率 ω 变大时，可以得到 η 变得较小，其中常数 ε 可以防止随着迭代的进行步长 η 逐渐减小且不会变为 0，可以保证不会终止训练。在迭代更新频率 ω 变小时，η 逐渐变大。

SSU-SGD 算法使用指数衰减平均，能够找到凸结构后快速收敛，为防止过拟合，在研究收敛率时，讨论额外误差（excess error）$J(\lambda) - \min_\lambda(\lambda)$ 在 k 次迭代后额外误差量级是 $o(\frac{1}{\sqrt{k}})$。

用 $E[\Delta\lambda^2]_{t-1}$ 代替步长 η 的得到 SSU-SGD 算法更新公式如式（6）（7）所示。

$$\Delta\lambda_{t+1} = -\frac{E[\Delta\lambda^2]_{t-1}}{E[\nabla^2]_t} \otimes \nabla_\lambda J(\lambda_t) \tag{6}$$

$$\lambda_{t+1} = \lambda_t + \Delta\lambda_{t+1} \tag{7}$$

其中：$E[\Delta\lambda^2]_t$ 和 $E[\nabla^2]_t$ 为均方根误差，分别如式（8）（9）所示：

$$E[\Delta\lambda^2]_t = \gamma E[\Delta\lambda^2]_{t-1} + (1-\gamma)\Delta\lambda_t^2 \tag{8}$$

$$E[\nabla^2]_t = \gamma E[\nabla^2]_{t-1} + (1-\gamma)\nabla_\lambda J(\lambda_t)^2 \tag{9}$$

γ 赋值为 0.9，此算法的优点就是可以避免手动调节步长，一般设置初始值为 0.01，让其在学习的过程中自己变化，同时可以避免 SGD 算法的训练过程中分母累积的和越来越大，保证在学习的后来阶段网络更新能力不会减弱。当应用于非凸函数训练神经网络时，学习轨迹在穿过不同的结构后到达一个局部是凸碗的区域，这种方法可以保证快速收敛，就像是初始化于该碗状结构的 SGD 算法。

下面是步长的更新流程，SGD 的步长自学习更新算法步长更新流程图如下所示。

输入：当前迭代次数 n_i。

初始化 $\eta_0 = 0.01$，$\varepsilon = 1e-8$

//初始化步长 η 和设置极小值常数

while $n_i < N$　　//判断当前迭代数是否小于总迭代数

if $n_i > n_0$　　//当前迭代次数大于设置的阈值

$\eta \leftarrow R^{(n_i/\omega)} * (\eta_0 / \gamma^{n/\omega_0}) + \varepsilon$　　//步长自学习法则 1

else　　//当前迭代次数小于设置的阈值

$\eta \leftarrow R^{(n_i/\omega)} * \eta_0 + \varepsilon$　　//步长自学习法则 2

输出 η。

说明：

R :每次更新的幅度

ω：迭代更新的频率

ω_0：迭代更新的初始频率

2　基于 SSU-SGD 的动态手持物体识别方法

2.1　动态手持物体识别的基准

将深度学习方法应用到动态手持物体识别[22]，连续目标识别不仅要考虑物体的形状、大小、位置、光线还要考虑物体的运动方向、运动轨迹等信息[23~25]。在迭代过程中需要训练的样本量巨大，而且训练样本所处的环境复杂多变，所以在每次迭代的过程中系统的整体性能和稳定性较差，识别率不高，因此提出 SSU-SGD 算法，根据特征出现的频率使相应参数可以自学出不同的步长，可以快速得到动态手持物体识别问题的最优解。

本文根据提出的 SSU-SGD 模型，结合深度学习方法给出三种适用于动态手持物体识别的基准，下面简单介绍一下三种步长自学习 SGD 的动态手持物体识别基准（benchmark）：

Benchmark1　在训练过程中自学习出不同的步长和迭代次数，训练分为 8 个批次，对第一批次的某一场景下的所有对象进行训练，将得到的分类结果用于调整后序的动态手持物体识别训练，由于在第一批次的训练中可以得到分类结果，连续的训练批次是对已知类的改进和巩固训练。

Benchmark2　在训练过程中自学习出不同的步长和迭代次数，训练分为 8 个批次，每一个批次都完成一次 8 个场景下的分类训练，不能将第一批次的训练结果用于后序训练，所以每一批次都是在未知类的基础上进行训练，连续的训练批次用于巩固训练。

Benchmark3　在训练过程中自学习出不同的步长和迭代次数，对每一批次都进行分类训练，每个训练批次不仅可以得到分类结果还可以达到对每个对象的不同场景下的识别要求，连续的训练批次是对未知类和对象的改进和巩固训练。

2.2　CORe50 数据集

本文采用 CORe50 数据集，该数据集共 50 个对象分成了 10 类用于实验，考虑物体的位置和光线，分别在 11 个不同的场景（8 个室内和 3 个室外）收集 300 个 RGB-D 帧的动态图像，相当于 50×11×300 帧图像用于训练。随机挑选 3 个场景（包括室内和室外）下的 50×3×300 帧图像用于测试，其余 8 个场景的 50×8×300 帧图像用于训练。如图 2-2 所示，包括插头适配器、移动电话、剪刀、灯泡、罐、眼镜、球、记号笔、杯子和遥控器等 10 类手持物品。

3　实验过程及结果

3.1　实验平台和方法

实验平台选择的是操作系统为 Linux Ubuntu16.04，选择的 GPU 型号 NVIDIA GTX1080Ti，主频为 11 GHz，可用显存大小为 11 GB。

本文在 SSU-SGD 算法的基础上提出适用于动态手持物体识别的三个基准，该基准可以根据特征出现的频率自学习出不同的步长，然后分别对每个基准用 Caffe 中的 Alexnet

网络和 VGG 网络训练。将 CORe50 数据集中 50*8*300 帧图像用于训练，50*3*300 帧图像用于测试。在动态手持物体识别策略中 naive 策略仅仅是显示当前批次的训练结果，而累积策略显示当前批次和前批次的训练结果，所以理论上累积策略的方法更有效，本文分别对 Alexnet 网络和 VGG 网络用 naive 策略和累积策略训练，以 Benchmark2（B2）为例，给出训练层级结构示意图如图 2 所示。



图 1 CORe50 数据集样本

Fig. 1 Samples of CORe50 dataset

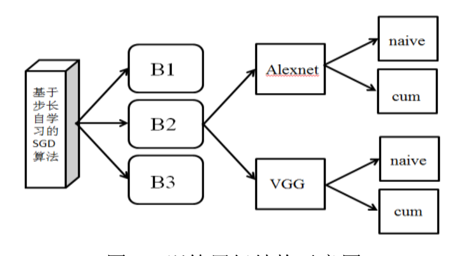


图 2 训练层级结构示意图

Fig. 2 Training hierarchy

3.2 实验结果与分析

本实验进行了 2000 次迭代，以预测精确度作为实验性能评价的指标，为了衡量本文所提方法的对动态 RGB-D 场景的识别性能，根据本文提出的动态手持物体识别的三个基准分别进行实验，得到图 3~5 所示的结果。

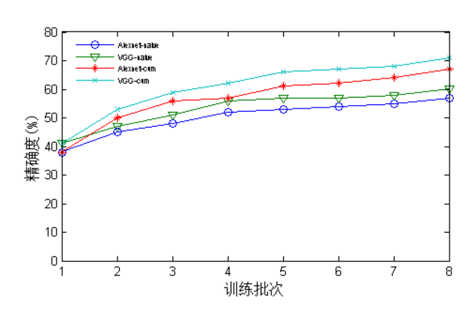


图 3 Benchmark1 在不同网络下的识别精确度

Fig. 3 Recognition accuracy of Benchmark 1 in different networks

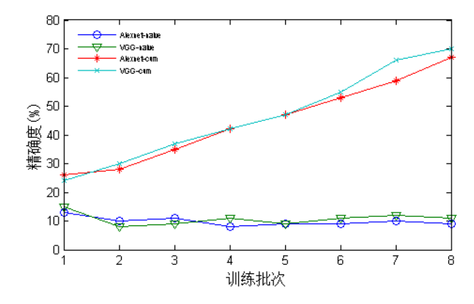


图 4 Benchmark2 在不同网络下的识别精确度

Fig. 4 Recognition accuracy of Benchmark 2 in different networks

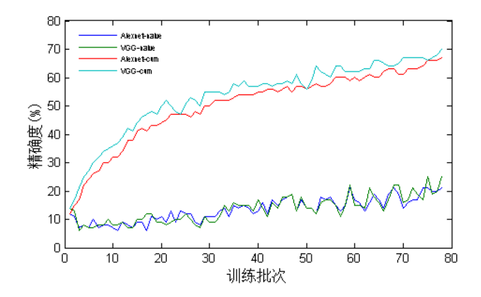


图 5 Benchmark3 在不同网络下的识别精确度

Fig. 5 Recognition accuracy of Benchmark3 in different networks

其中 Benchmark1 下 Alexnet 网络和 VGG 网络得到的精确度较高，两种识别策略表现良好，在 Benchmark2 和 Benchmark3 中不同网络下的 naive 策略遭到了灾难性遗忘，而累积策略表现良好，Benchmark3 是最接近实际生活的动态手持物体识别基准，将训练的 batches 改为 78，可以更加精确地看到测试结果的变化。综合比较图 3~5 可以看出 VGG 网络的测试结果都达到 70%左右，优于 Alexnet 网络，尤其在 Benchmark2 和 Benchmark3 中 naive 策略变得完全不可用，通过不同基准、不同网络的比较，可以看出 VGG 网络中的累积策略可以较好的满足动态手持物体识别的要求。

3.3 测试结果的验证

为验证本文方法的有效性，将基于 SSU-SGD 算法的动态手持物体识别的三个基准与传统的识别基准 Mid-CNN from scratch 和 Mid-CNN+SVM 进行对比。对 CORe50 数据集进行训练测试，得到验证结果如表 1 所示。

表 1 不同基准下的识别精确度

Table 1 Recognition accuracy under different benchmarks

accuracy/%	Alexnet		VGG	
	Naive	Cum	Naive	Cum
Mid-CNNfrom scratch	37.35	38.33	48.28	53.73
Mid-CNN+SVM	51.35	59.03	61.08	68.03
Benchmark1	**57.38**	**67.64**	**60.41**	**71.68**
Benchmark2	**9.13**	**67.26**	**11.15**	**70.66**
Benchmark3	**21.16**	**67.56**	**25.37**	**70.14**

表 1 给出了 Mid-CNN from scratch 和 Mid-CNN+SVM

基准下的识别精确度和本文提出的基于步长自学习更新的SGD 动态手持物体识别基准相比较，可以看出本文提出的基于 SSU-SGD 算法的动态基准可以更好的应用于动态手持物体识别，得到更高的识别精确度。

接下来给出 SSU-SGD 算法和传统 SGD 算法的识别精确度，如表 2 所示。

表 2 SSU-SGD 算法与 SGD 算法的实验对比

Table 2 Experimental comparison of SSU-SGD algorithm and SGD algorithm

accuracy/%		Alexnet		VGG	
		Naive	Cum	Naive	Cum
SGD	Benchmark1	57.02	65.64	59.35	68.23
	Benchmark2	9.12	67.04	11.23	68.58
	Benchmark3	20.68	65.35	25.23	69.02
SSU-SGD	**Benchmark1**	**57.38**	**67.64**	**60.41**	**71.68**
	Benchmark2	**9.13**	**67.26**	**11.15**	**70.66**
	Benchmark3	**21.16**	**67.56**	**25.37**	**70.14**

表 2 给出了 SSU-SGD 算法和传统 SGD 算法的识别精确度，可以看出本文提出的基于 SSU-SGD 算法的基准可以更好的应用于动态手持物体识别，得到更高的识别精确度。

4 结束语

本文在研究动态手持物体识别的基础上改进了基于 SSU-SGD 的动态手持物体识别基准，使用 Alexnet 网络和 VGG 网络同时训练，提高了训练速度和动态手持物体识别的精确度，SSU-SGD 算法可以保证用较少的数据与使用全部数据的训练结果一致，大大提高训练的速度，同时保证产生较小的训练震荡基于 SSU-SGD 算法提出的动态手持物体识别的三个基准相比较传统的 Mid-CNN from scratch 和 Mid-CNN+SVM 基准可以较好满足动态手持物体识别的要求。由于受到现有数据库的限制，本文没有对复杂环境下的动态物体识别测试，在后面的工作中将在构建复杂场景动态样本库的基础上测试提出的方法，进而提高算法的可靠性，并将更加优化的算法进一步应用到视频目标识别中。

参考文献：

[1] Bilen H, Fernando B, Gavves E, *et al.* Dynamic image networks for action recognition [C]//Computer Vision and Pattern Recognition. IEEE, 2016: 3034-3042.

[2] Shi Wei. Design of web interface based on visual information communication [J]. Revista De La Facultad De Ingenieria, 2017, 32 (11): 679-684.

[3] Zhang Ling, Jiang Yi, Li Y, *et al.* Adaptive maneuvering target tracking with 2-HFSWR multisensor surveillance system [J]. IEEE Aerospace & Electronic Systems Magazine, 2018, 32 (12): 70-76.

[4] 杜兰，刘彬，王燕，等. 基于卷积神经网络的 SAR 图像目标检测算法 [J]. 电子与信息学报, 2016, 38(12): 3018-3025. (Du Lan, Liu Bin, Wang Yan, *et al.* SAR image target detection algorithm based on convolution neural network [J]. Journal of electronics and information, 2016, 38 (12): 3018-3025.)

[5] Schwarz M, Schulz H, Behnke S. RGB-D object recognition and pose estimation based on pre-trained convolutional neural network features [C]//Proc of IEEE International Conference on Robotics and Automation. IEEE, 2015: 1329-1335.

[6] Luo Zelun, Peng Boya and Huang De-An, *et al.* Unsupervised learning of long-term motion dynamics for videos [J]. Computer Vision and Pattern Recognition, 2017: 7101-7110.

[7] Borji A, Izadi S, Itti L. iLab-20M: A large-scale controlled object dataset to investigate deep learning [C]// Computer Vision and Pattern Recognition. IEEE, 2016: 2221-2230.

[8] Liu Zhenan, Yan Tingrong, Zhang Rui. Dynamic image recognition in static background [J]. Computer technology and development, 2001, 11 (1): 52-53.

[9] Horiguchi S, Amano S, Ogawa M, *et al.* Personalized classifier for food image recognition [J]. IEEE Trans on Multimedia, 2018, P(99): 1-1.

[10] Parthasarathy S, Rose R C. System and method for mobile automatic speech recognition [J]. Journal of the Acoustical Society of America, 2018, 124(6): 3373-3373.

[11] 褚晓敏，朱巧明，周国栋. 自然语言处理中的篇章主次关系研究 [J]. 计算机学报, 2017, 40(4): 842-860. (Zhu Xiaomin, Zhu Qiaoming and Zhou Guodong. Research on the relationship between major and subordinate in Natural Language Processing [J]. Journal of computer science, 2017, 40(4): 842-860.)

[12] Baldi P, Sadowski P, Whiteson D. Searching for exotic particles in high-energy physics with deep learning [J]. Nature Communications, 2014, 5(5): 4308.

[13] Zhou Jian, Troyanskaya Olga G. Predicting effects of noncoding variants with deep learning–based sequence model [J]. Nature Methods, 2015, 12(10): 931-934.

[14] Liu Weiwei. Video face detection based on deep learning [J]. Wireless Personal Communications, 2018(12): 1-16.

[15] Yuan Zhengwu, Zhang Jun. Feature extraction and image retrieval based on AlexNet [C]//Proc of the 8th International Conference on Digital Image Processing. 2016: 100330E.

[16] Verschuere B, Sophia V G G, Waldorp L, *et al.* What features of psychopathy might be central? A network analysis of the Psychopathy Checklist-Revised (PCL-R) in three large samples [J]. Journal of Abnormal Psychology, 2018, 127(1) .

[17] Sa C D, Feldman M and Re C, *et al.* Understanding and optimizing asynchronous low-precision stochastic gradient descent [C]// Proc of ACM/IEEE, International Symposium on Computer Architecture. 2017: 561-574.

[18] 邓卫钊. 随机梯度下降和对偶坐标下降算法的研究与应用 [D]. 秦皇岛: 燕山大学, 2016. (Deng Weizhao. Research and application of stochastic gradient descent and dual coordinate descent algorithm [D]. Qinhuangdao: Yanshan University, 2016.)

[19] 金钊. 基于改进随机梯度下降算法的 SVM [D]. 保定: 河北大学, 2017. (Jin Zhao. SVM based on improved stochastic gradient descent algorithm [D]. Baoding: Hebei University, 2017.)

[20] Klein S, Pluim J P, Staring M, *et al.* Adaptive stochastic gradient descent optimisation for image registration [J]. International Journal of Computer Vision, 2009, 81(3): 227.

[21] Gao Yuan, Ma Jiayi, Alan L. Yuile, *et al.* Semi-supervised sparse representation based classification for face recognition with Insufficient Labeled samples [J]. IEEE Trans on Image Processing A Publication of the IEEE Signal Processing Society, 2017, 26(5): 2545.

[22] Liu Daqi, Yue Shigang. Event-driven continuous STDP learning with deep structure for visual pattern recognition [J]. IEEE Trans on Cybernetics, 2018: 1-14

[23] Sobhani B, Paolini E and Giorgetti A, *et al.* Target tracking for UWB multistatic radar sensor networks [J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 8(1): 125-136.

[24] 王肖锋，张明路，刘军. 基于增量式双向主成分分析 的机器人感知学习方法研究 [J]. 电子与信息学报，2018，40(3)：618-625. (Wang Xiao Feng, Zhang Minglu, Liu Jun. Research on robot perceptual learning based on incremental two-way principal component analysis [J]. Journal of electronic and information, 2018, 40(3): 618-625.)

[25] Imran S, Ko Y B. A continuous object Boundary detection and tracking scheme for Failure-Prone sensor networks [J]. Sensors, 2017, 17(2): 360-361.